Using Markov chain Monte Carlo methods for estimating parameters with gravitational radiation data.

Nelson Christensen[1] and Renate Meyer[2*]

[1] *Physics and Astronomy, Carleton College, Northfield, Minnesota, 55057, USA*

[2] *Department of Statistics, The University of Auckland, Auckland, New Zealand*

(February 27, 2001)

Abstract

We present a Bayesian approach to the problem of determining parameters for coalescing binary systems observed with laser interferometric detectors. By applying a Markov Chain Monte Carlo (MCMC) algorithm, specifically the Gibbs sampler, we demonstrate the potential that MCMC techniques may hold for the computation of posterior distributions of parameters of the binary system that created the gravity radiation signal. We describe the use of the Gibbs sampler method, and present examples whereby signals are detected and analyzed from within noisy data.

04.80.Nn, 02.70.Lq, 06.20.Dq

I. INTRODUCTION

A number of collaborations around the world will be operating laser interferometric gravitation radiation antennas within the next few years. In the United States the Laser Interferometric Gravitational Wave Observatory (LIGO) is soon to be operational, with 4 km arm length interferometers in Hanford, Washington, and Livingston, Louisiana [1]. A similar French-Italian detector will be built in Europe (VIRGO) [2,3].

Coalescing binaries containing neutron stars (NS) or black holes (BH) are likely to be the cleanest and most promising source of detectable radiation [4]. Ultimately the LIGO-VIRGO network may observe binaries out to a distance of 2 Gpc [5]. The detection of coalescing binary events will provide physicists with extremely useful cosmological information. Initially Schutz [6] noted that a detected signal contains enough information to decipher the absolute distance to the system, and hence the determination of the Hubble constant would be achieved through the observed distribution of several binaries. Subsequent work [7] indicates that the uncertainty in the measured distance can be comparable to the distance itself, but important cosmological tests will still be possible through the observation of numerous mergers [8].

In addition to the cosmological importance, accurate parameter estimation in the observed coalescing binaries will provide a host of information of great physical significance. Observation of the time of tidal disruption of an NS - NS binary system may permit a determination of the NS radii and information on the NS equation of state [9]. The characteristics of radiation in the post-Newtonian regime will provide insight into highly non-linear general relativistic effects [7,10,11]. The formation of a BH at the end of a NS-NS coalescence, or the merger of two BHs, will produce gravitational radiation as the system decays to a Kerr BH; this is an extremely interesting radiation production regime [10,11].

Application of Bayes' theorem is well suited to astrophysical observations [12]. The Bayesian versus frequentist approaches to gravitational radiation data analysis are well presented by [13]. Parameter estimation from the gravity wave signals of coalescing compact binaries provides an important application of Bayesian methods [5,7,14,15]. Difficulties with the calculation of Bayesian posterior distributions have been overcome by the rapid development of Markov Chain Monte Carlo (MCMC) methods in the last decade (see [16] for an introduction). Although the initial MCMC algorithm dates back to [17], the enormous potential that MCMC methods might hold for Bayesian posterior computations remained largely unrecognized within the statistical community until the seminal paper by Geman and Geman [18] in the context of digital image analysis. Since then, MCMC methods have had a huge impact on many areas of applied statistics. It has now become practical to apply Bayesian methods to complex problems. Thus, we expect a similar effect on gravitational wave data analysis.

The initial goal of our research effort, presented in this paper, is to demonstrate the usefulness of MCMC techniques for estimating parameters from coalescing binary signals detected by laser interferometric antennas. The Gibbs sampler [16] is one of the simpler MCMC techniques, and we use it as a starting point for our investigation primarily because there is readily available software [19]. Our study of gravity wave signals is conducted to 2.5 post-Newtonian (PN) order. The signals depend on five independent parameters; the masses of the two compact objects, the amplitude of the detected signal, the coalescence time and the phase of the signal at coalescence. The Bayesian techniques we employ will not only give point estimates of these parameters, but also produce their complete posterior probability distribution that can be employed to summarize the uncertainty of parameter estimates through posterior credibility intervals, for instance. In contrast to frequentist confidence intervals, these do not rely on large sample asymptotics and have a simple, natural interpretation.

The paper is organized as follows: In Section II we briefly review Bayesian inference and describe the MCMC simulation technique we use, specifically the Gibbs sampler and software for its implementation. In Section III we present two examples where we use our MCMC approach to identify the parameters which created the signal that is buried in synthesized LIGO noise. In Section IV we analyze a number of issues that will effect the efficiency and calculational time of a MCMC approach to the coalescing binary parameter estimation problem. We conclude with a discussion of our results and the direction of future efforts in Section V.

II. BAYESIAN INFERENCE AND POSTERIOR COMPUTATION

We briefly review the Bayesian approach to parameter estimation. Let us assume the data consists of n observations, $\mathbf{z} = (z_1, \ldots, z_n)$, with joint PDF denoted by $p(\mathbf{z}|\boldsymbol{\theta})$ conditional on unobserved parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. The PDF $p(\mathbf{z}|\boldsymbol{\theta})$ is usually referred to as the *likelihood* and regarded as a function of $\boldsymbol{\theta}$. In contrast to the frequentist approach where $\boldsymbol{\theta}$ is regarded as fixed but unknown, the Bayesian approach treats $\boldsymbol{\theta}$ as a random variable with a probability distribution that reflects the researcher's uncertainty about the parameters. Bayesian inference requires the specification of a prior PDF for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$, that should take all information into account that is known about $\boldsymbol{\theta}$ before observing the data. All information about $\boldsymbol{\theta}$ that stems from the experiment should be contained in the likelihood. Bayesian inference then answers the question: "How should the data $\mathbf{z}$ change the researcher's knowledge about $\boldsymbol{\theta}$?" Via an application of Bayes' theorem, by conditioning on the known observations, this post-experimental knowledge about $\boldsymbol{\theta}$ is expressed through the *posterior* PDF

$$p(\boldsymbol{\theta}|\mathbf{z}) = \frac{p(\boldsymbol{\theta})p(\mathbf{z}|\boldsymbol{\theta})}{m(\mathbf{z})} \propto p(\boldsymbol{\theta})p(\mathbf{z}|\boldsymbol{\theta}) \tag{1}$$

where $m(\mathbf{z}) = \int p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the marginal PDF of $\mathbf{z}$ which can be regarded as a normalizing constant as it is independent of $\boldsymbol{\theta}$. The posterior PDF is thus proportional to the product of prior and likelihood.

The standard Bayesian point estimate of a single parameter, say θ_i, is the posterior mean

$$\hat{\theta}_i = \int \theta_i p(\theta_i|\mathbf{z}) d\theta_i \tag{2}$$

where

$$p(\theta_i|\mathbf{z}) = \int \ldots \int p(\boldsymbol{\theta}|\mathbf{z}) d\theta_1 \ldots d\theta_{i-1} d\theta_{i+1} \ldots d\theta_d. \tag{3}$$

is the marginal posterior PDF obtained by integrating the joint posterior PDF over all other components of $\boldsymbol{\theta}$ except θ_i. A measure of the uncertainty of this estimate is the posterior standard deviation or a 95% credibility interval that contains the parameter θ_i with 95% probability, its lower and upper bound being specified by the 2.5% and 97.5% percentile of $p(\theta_i|\mathbf{z})$, respectively. Alternatives to the posterior mean are the posterior mode, aka maximum a posteriori estimate (MAP), and the more robust posterior median.

As seen from equations (2) and (3), the calculation of posterior means requires d-dimensional integration, one of the main issues that has made the application of Bayesian inference so difficult in the past. This hurdle has been overcome by the great advances in simulation-based integration techniques, so called MCMC methods [16,20]. In MCMC, a Markov chain is constructed with the joint posterior as its equilibrium distribution. Thus, after running the Markov chain for a certain "burn-in" period, one obtains (correlated) samples from the limiting distribution, provided that the Markov chain has reached convergence. One popular construction principle is the Gibbs sampler, a specific MCMC method that samples iteratively from each of the univariate full *conditional* posterior distributions

$$p(\theta_i|\mathbf{z}, \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_d). \tag{4}$$

Given an arbitrary set of starting values $\theta_1^{(0)}, \ldots, \theta_d^{(0)}$ the algorithm proceeds as follows:

$$\begin{aligned} &\text{simulate } \theta_1^{(1)} \sim p(\theta_1|\mathbf{z}, \theta_2^{(0)}, \ldots, \theta_d^{(0)}) \\ &\text{simulate } \theta_2^{(1)} \sim p(\theta_2|\mathbf{z}, \theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_d^{(0)}) \\ &\qquad\qquad \vdots \\ &\text{simulate } \theta_d^{(1)} \sim p(\theta_d|\mathbf{z}, \theta_1^{(1)} \ldots, \theta_{d-1}^{(1)}) \end{aligned} \tag{5}$$

and yields $\boldsymbol{\theta}^{(m)} = (\theta_1^{(m)}, \ldots, \theta_d^{(m)})$ after m such cycles. This defines a Markov chain that converges to the joint posterior as its equilibrium distribution [16]. Consequently, if all the full conditional posterior distributions are available, all that is required is sampling iteratively from these. Thereby, the problem of sampling from an d-variate PDF is reduced to sampling from d univariate PDFs.

In many applications where the prior PDF is conjugate to the likelihood, the full conditionals in fact reduce analytically to closed-form PDFs and we can use highly efficient special purpose Monte Carlo methods for generating from these (see e.g. [21]). In general, however, we need a fast and efficient black-box method to sample from an arbitrarily complex full conditional posterior distribution in each cyclic step of the Gibbs sampler. Such an all-purpose algorithm, so-called *adaptive rejection sampling* (ARS) was developed by Gilks and Wild [22] for the rich class of distributions with *log-concave* densities. We can use a recently developed "Metropolized" version of adaptive rejection sampling (ARMS) for non-logconcave distributions [17,23]. C-subroutines of ARS and ARMS are available [23] and can thus be tailored to the full conditional posteriors of the problem at hand.

Significant progress has been made in facilitating the routine implementation of the Gibbs sampler with the help of BUGS (Bayesian Inference Using Gibbs Sampling), a recently developed software package [19] by the Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, England. BUGS samples from the joint posterior distribution by using the Gibbs sampler. For reviews on BUGS the reader is referred to [24–26]. BUGS is available free of charge from

`http://www.mrc-bsu.cam.ac.uk/bugs/Welcome.html`

BUGS can handle the two main tasks necessary for implementation of the Gibbs sampler. These tasks are to i) construct and ii) to sample from the full conditional posterior densities. Only the prior and sampling distributions for unobservables and observables, respectively, have to be specified in a BUGS program. The tedious task of constructing the full conditionals is automated by BUGS using directed acyclic graphs [27]. Sophisticated routines such as adaptive rejection sampling to sample from log-concave full conditionals and MH algorithms based on slice sampling to sample from non-logconcave full conditional densities have been implemented in BUGS and are continuously being refined. Furthermore, various methods to assess convergence, i.e. methods used for establishing whether an MCMC algorithm has converged and whether its output can be regarded as samples from the target distribution of the Markov chain, have been developed and implemented in CODA [28]. CODA is a menu-driven collection of SPLUS functions for analyzing the output obtained by BUGS. Besides trace plots and the usual tests for convergence, CODA calculates statistical summaries of the posterior distributions and kernel density estimates. CODA is being maintained and distributed by the same research group responsible for BUGS.

III. EXAMPLES OF COALESCING BINARY SIGNALS DETECTED BY A LASER INTERFEROMETER

Our initial goal, presented in this paper, is to demonstrate the usefulness of MCMC techniques for estimating parameters from coalescing binary signals detected by laser interferometric antennas. We generated the coalescing binary gravity wave signals to 2.5 PN order in both the time [29,30] and frequency [31] domains. Noise that simulates the LIGO II environment was synthesized in the time domain using software from Finn and Daw [32]. The power spectral density of the noise was calculated from long-time noise signals generated by [32]. The 2.5 PN frequency domain signals also served as the *templates* for our extraction of the event parameters.

The total output registered by the detector, $z\,(t)$, is the sum of the gravity wave signal, $s(t,\boldsymbol{\theta})$, that depends on unknown parameters $\boldsymbol{\theta}$, and the noise $n(t)$, namely $z\,(t) = s\,(t,\boldsymbol{\theta}) + n\,(t)$ for $t \in [0, t_u]$. We assume that the noise is Gaussian with mean zero and known one-sided power spectral density $S_n(f)$. The signal-to-noise ratio (SNR) of the *detected* signal is $SNR = \sqrt{2\int_{-\infty}^{\infty} \frac{\tilde{s}(f)\tilde{s}(f)}{S_n(f)} df}$ with $\tilde{s}\,(f) = \int_{-\infty}^{\infty} s\,(t)\, e^{2\pi i f t} dt$ the Fourier transform of the function $s(t)$ [8]. The *likelihood* is given by

$$p(\mathbf{z}|\boldsymbol{\theta}) = K \exp\left[2\left\langle z, s(\boldsymbol{\theta})\right\rangle - \left\langle s(\boldsymbol{\theta}), s(\boldsymbol{\theta})\right\rangle\right] \tag{6}$$

where K is a constant and $\langle a, b\rangle = \int_{-\infty}^{\infty} df \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}$ the inner product of two functions a, b [5].

Since $\tilde{z}(f) = \tilde{z}^*(-f)$ for real signals, we can express the likelihood as

$$p(\mathbf{z}|\boldsymbol{\theta}) = K \exp\left[-2\int_0^\infty df \frac{Re\left\{(\tilde{z}(f)-\tilde{s}(f,\boldsymbol{\theta}))(\tilde{z}(f)-\tilde{s}(f,\boldsymbol{\theta}))^*\right\}}{S_n(f)}\right]. \tag{7}$$

For discretized data the likelihood takes the form

$$p(\mathbf{z}|\boldsymbol{\theta}) = K * \exp\left[-2\sum_{i=i_l}^{i_u} \frac{Re\left(\tilde{z}(i*\Delta f)-\tilde{s}(i*\Delta f,\boldsymbol{\theta})\right)\left(\tilde{z}(i*\Delta f)-\tilde{s}(i*\Delta f,\boldsymbol{\theta})\right)^*}{S_n\left(i*\Delta f\right)}\right] \tag{8}$$

where $i_l * \Delta f$ and $i_u * \Delta f$ correspond to the lower and upper limits of the frequency range examined and Δf is the resolution of the discretized domain data.

As detailed in [31], the signal depends on five parameters, namely the masses m_1, m_2 of the two compact objects, the coalescence time t_c, the phase of the wave φ_0, and the amplitude N through

$$\tilde{s}(f,\boldsymbol{\theta}) = N e^{i\varphi_0} f^{-7/6} e^{i(\psi(f)+2\pi f t_c)} \tag{9}$$

with $\boldsymbol{\theta} = (m_1, m_2, t_c, \varphi_0, N)$ and

$$\psi(f) = \sum_{i=1}^{5} a_i \varsigma_i(f) \tag{10}$$

where

$$\begin{aligned} a_1 &= \frac{3}{128\eta} q^{-5/3}, \\ a_2 &= \frac{1}{384\eta}\left(\frac{3715}{84} + 55\eta\right) q^{-1}, \\ a_3 &= \frac{-1}{128\eta} 48\pi q^{-2/3}, \\ a_4 &= \frac{3}{128\eta}\left(\frac{15293365}{508032} + \frac{27145}{504}\eta + \frac{3085}{72}\eta^2\right) q^{-1/3}, \\ a_5 &= \frac{\pi}{128\eta}\left(\frac{38645}{252} + 5\eta\right), \end{aligned} \tag{11}$$

and $\varsigma_1 = f^{-5/3}$, $\varsigma_2(f) = f^{-1}$, $\varsigma_3(f) = f^{-2/3}$, $\varsigma_4(f) = f^{-1/3}$, $\varsigma_5(f) = \ln(f)$, the total mass $m_t = m_1 + m_2$, $q = \pi G m_t / c^3$, and the mass ratio $\eta = m_1 m_2 / m_t^2$.

The *templates* for the gravity wave signal, $\tilde{s}(f, \boldsymbol{\theta})$, were always generated in the frequency domain to 2.5PN order according to the formulas given in [31] because the entire likelihood computation is also done entirely in the frequency domain. Examples of our Gibbs sampler code for use with BUGS can be obtained electronically [33].

A. Results for Signals F in the Frequency Domain

The published 2.5 PN order time domain templates are not quite the exact Fourier transform of the 2.5 PN frequency domain templates, hence we felt the necessity to test our method with signals we generated in both time or frequency space. For signals generated in the frequency domain we used published 2.5 PN order results [31]. For this example we chose the masses of the compact objects to be $m_1 = 1.4M_\odot$ and $m_2 = 3.5M_\odot$, the coalescence time was $t_c = 1ms$, and the phase of the wave of $\varphi_0 = 0.123$. The amplitude of the wave was adjusted to create appropriate SNR values. The noise was generated in the time domain [32], and subsequently transformed to the frequency domain.

We assumed noninformative *a priori* distributions for all of our parameters; namely a uniform distribution on $0.3M_\odot$ to $12M_\odot$ for each of the two compact objects (m_1 and m_2) and a uniform distribution on 0 to $50ms$ for t_c. The gravity wave phase will lie between and $-\pi$ to π for φ_0, but we assume a uniform a priori distribution between -2π to 2π so that the converged chain can more easily sample its region of interest. The amplitude of the incoming gravity wave has a dependence of $\tilde{h}\,(f) = N\eta^{1/2}m_t^{5/6}f^{-7/6}$. Our simulated signal thereby has $m_t = 4.9M_\odot$ and $\eta = 0.2041$. We used a uniform *a priori* for the amplitude term N on the interval -10^{-19} to 10^{-19}. This was subsequently renormalized with a factor of 10^{25} for computational reasons yielding a uniform *a priori* distribution for N on the interval -10^6 to 10^6. We assumed that the compact objects had no spin, and hence these parameters were not included in this study.

Unique gravity wave signals do not depend on m_1 and m_2, but instead on the total mass m_t and the mass ratio η. Hence it is the posterior probability distribution functions of m_t

and η that provide the best information about the system. The posterior distributions of m_1 and m_2 from the MCMC are unnecessarily wide due to oscillations of each chain between the two possible values. Estimates and statistical properties of m_1 and m_2 must be inferred from the distributions for m_t and η.

In our initial implementation of the Gibbs sampler it was found that it takes a prohibitively long time for the chain to *burn in* and sample from the correct posterior distribution. Instead, an efficient procedure that allowed the chain to more efficiently explore the phase space was one that is analogous to *simulated annealing* [34]. In this procedure we use a likelihood of the form $L = K \exp\left[2\left\langle z,s\right\rangle - \left\langle s,s\right\rangle\right]$ with $\left\langle z,s\right\rangle = \int_{-\infty}^{\infty} df \frac{\tilde{z}(f)\tilde{s}^*(f)}{T*S_n(f)}$, $\left\langle s,s\right\rangle = \int_{-\infty}^{\infty} df \frac{\tilde{s}(f)\tilde{s}^*(f)}{T*S_n(f)}$. The auxiliary variable T is a *pseudo-temperature*. If T is chosen large, $T >> 1$, we have *heating* and the MCMC does not get trapped in particular regions of phase space for too long. It essentially corresponds to increasing the variance of the posterior distribution to allow for wider jumps. Thus, the chain can reach all regions of the state space. In our study we typically start with T˜500 and allow the chain to "burn-in" and find equilibrium. For each value of T the mean values for the parameters are computed and used as the starting values for the next chain with its reduced value of T. The T term can be quickly brought to $T = 1$ and the final kernel densities generated. Combining simulated annealing with MCMC samplers has been demonstrated to improve the efficiency of chains that mix poorly in their phase space [35].

When we analyzed the data in our MCMC program we had a frequency resolution of $\Delta f = 1Hz$ and only utilized the frequency range $30Hz - 730Hz$; this range was based on reasonable assumptions of LIGO performance. Reducing the upper frequency limit does not greatly effect the performance of the MCMC results; this makes sense because for a coalescing binary signal most of the power of the signal is at the lower frequencies. The largest amplitude signal may happen at a high frequency, but the binary emits more cycles at lower frequencies. Although we have not studied this issue in depth, it appears that good MCMC performance will result even if one only includes frequencies up to ˜$200Hz$.

In Figs. 1-3 we present the results of this part of our investigation. We see that for large

$SNRs$ we can accurately predict the parameters producing the signal. At the lower $SNRs$ it takes the chain much longer to burn-in and find the correct parameters, and when the SNR is too low the chain will not converge and no useful information will be discerned. The results for $SNR = 4.4$ are further illustrated in Fig. 4, where we can observe the kernel densities for the parameters generated from the Gibbs sampler.

Fig. 5 displays the operation of the Gibbs sampler, whereby the trace plots for the parameters are displayed. This is the result for $SNR = 4.4$ when we have set the annealing temperature to $T = 100$. One can see how the chain *burns-in* and achieves convergence. We would let the chain run for ˜10,000 iterations, compute the mean values for the parameters (excluding points prior to *burn-in*), and use the calculated mean parameter values as initial conditions for a new Markov chain with a diminished annealing temperature. Once we have a $T = 1$ then 10,000 iterations typically produce an adequate and informative kernel density.

Extensive convergence diagnostics were calculated for all of the parameters using the CODA software [28]. All chains passed the Heideberger - Welch stationary test. The Raferty - Lewis convergence diagnostics confirmed the thinning and burn-in were sufficient. Lags and autocorrelations within each chain were reasonably low. Geweke Z-scores were low for all parameters. These convergence diagnostics are described in [28] (and references therein).

Continuing the use of the $SNR = 4.4$ example, we can decipher the masses of the individual compact objects. The result of the Gibbs sampler can give us estimates of these parameters by using simple summary statistics like the sample average and empirical percentiles. These yield posterior means of $m_t = 4.891 M_\odot$ and $\eta = 0.2048$, plus 2.5 to 97.5 percentile ranges of $m_t = (4.882 \text{ to } 4.898)\, M_\odot$ and $\eta = 0.2043$ to 0.2055. This then implies, compact object masses of $m_1 = (3.485 \pm 0.01)\, M_\odot$ and $m_2 = (1.406 \pm 0.008)\, M_\odot$.

B. Results for Signals Generated in the Time Domain

When we generated signals in the time domain the parameters were also chosen arbitrarily. The generated signals were made to 2.5 PN [29,30]. The masses of the two compact object were $8M_\odot$ and $9M_\odot$, the angle of inclination of the orbit was $\iota = \pi/4$, and the orientation of the gravity wave source with respect to the laser interferometer was $\varphi = 2.2$, $\theta = 1.1$, and $\psi = 3.3$ (all in radians). In order to adjust the signal to noise ratio we effectively changed the source to detector distance. The interferometer noise was also computer generated [32], and the signal and noise were summed together. We created signals of $32s$ length, with 16384 data points per second. The temporal signal was Fourier transformed to the frequency domain.

Templates for the likelihood were again 2.5 PN [31]. The results below were based on a MCMC investigation that ranged from $30Hz - 130Hz$, with frequency resolution of $\Delta f = 1Hz$. Increasing the upper frequency did not affect the results, but only slowed the calculation. This is again consistent with the fact that most signal power is at the lower frequency. The choice of an upper frequency for the MCMC will effect the speed of the calculation and ultimately its ability to accurately estimate parameters; this is a topic we are currently investigating and will be the subject of a future publication.

In Figs. 6-8 we present the results of this part of our investigation. We see that for large $SNRs$ we can again accurately predict the parameters producing the signal. The results for $SNR = 4.4$ are presented in Fig. 9, where we can observe the kernel densities for the parameters that we generated from the Gibbs sampler. Fig. 10 displays the operation of the Gibbs sampler, whereby the trace plots for the parameters are displayed. This is the result for $SNR = 4.4$ when we have set the annealing temperature to $T = 100$. One can see how the chain *burns-in* and converges.

IV. ANALYSIS ISSUES

There are a host of topics that need to be addressed before one could say that MCMC techniques will be truly applicable and useful with LIGO data. However, we have demonstrated that the Gibbs sampler does have potential usefulness, and can successfully find the signal and make statistical statements about the parameters. Numerous issues pertaining to MCMC use with LIGO data are discussed below.

The speed at which the MCMC calculation runs on computers is a concern of paramount importance. Numerous issues influence the speed of the calculation. For the study presented in this paper, we would generate $32s$ of data in the time domain. The data corresponded to a sampling rate of $16384Hz$. Due to the character of the signal source and LIGO noise we only considered signal frequencies above $30Hz$.

The choice of an upper frequency limit can significantly influence the speed of the MCMC program. However, one can not arbitrarily reduce the upper frequency too much or the ability to estimate parameters will degrade. If our templates only consider the orbital parameters of the binary system, and not the ringdown of the newly formed black hole then the largest useful frequency will correspond to twice the instantaneous orbital frequency of the last stable orbit before free-fall, $f = c^3 / \left(6^{3/2} \pi G m_t\right)$. In this study we only considered compact objects with masses between $0.3M_\odot$ and $10M_\odot$. The frequency could therefore range from as large $7300Hz$ for two $0.3M_\odot$ objects, to $1570Hz$ for two $1.4M_\odot$ neutron stars, to $220Hz$ for two $10M_\odot$ black holes. One should also remember that the Fourier transform of the signal falls off like $|h(f)| \propto f^{-7/6}$. Consequently the power of the signal is dominated by low frequencies; the binary spends more time emitting relatively low amplitude gravity waves at lower frequency as opposed to a shorter time producing larger amplitude signals at higher frequencies. Establishing an effective upper frequency choice will depend on the decision for the *a priori* distribution of the masses. It will also depend on analyses that investigate the behavior of the MCMC as the upper frequency varies. We anticipate that effective parameter estimation studies of events will vary the upper frequency and search for a convergence in behavior. As the upper frequency of the data is diminished the speed of the calculation increases.

The frequency resolution of the data is another aspect that will affect program speed and parameter estimation ability. For example, with $32s$ of data the Fourier transform points are separated by $1/32\ Hz$. We found this precise frequency resolution to be unnecessary. We increased the program speed without diminishing parameter estimation ability by decreasing the resolution to $1Hz$. All the results presented in this paper utilized a frequency resolution of $1Hz$. However, we feel that a detailed study of the coupling of frequency resolution with computational speed and effectiveness will be necessary.

The *simulated annealing* procedure we used to adequately sample the parameter space is another topic where studies will be needed in order to optimize the speed of the routine. For the variables we used in our program, either in terms of the compact masses m_1 and m_2 or the total mass m_t and the mass ratio η, it was necessary to introduce an *pseudo-temperature* T into the likelihood to initially *burn-in* the chain and converge to the correct parameters. Without T it would take a prohibitively long time for the chain to converge; the incorporation of simulated annealing into MCMC techniques has been demonstrated to decrease the burn-in time for the chain [35]. In the studies presented here we would typically start with $T\tilde{}500$, and it could take anywhere from 10^3 to 2×10^4 cycles for the chain to reach convergence. Smaller SNR events took longer to *burn-in*. The chain would run for about 10^4 after burn-in, and the mean values for the parameters were computed and used as the starting values for the next chain with its reduced value of T. The value of T would be decreased by an order of magnitude, so that a typical procedure would involve runs with $= 500$, 50, 5 and then 1. This cooling schedule is definitely not optimized for speed and efficiency. The use of better coordinates [36] may provide a more well behaved parameter space, which in turn could help the chain mix more efficiently and reduce the time needed for simulated annealing. This will be investigated.

All of the calculations presented in this paper were conducted on a 500 MHz pc. When the frequency span of the study extended from $30Hz$ to $730Hz$, with $1Hz$ resolution, 10^4 cycles of the MCMC took 1.5 hours. We often needed about 2.5×10^4 cycles to generate good kernel densities. When the frequency span extended from $30Hz$ to $130Hz$, with $1Hz$ resolution, 10^4 cycles of the MCMC typically took approximately 10 minutes.

V. DISCUSSION

In this paper we have demonstrated that the Gibbs sampler can be used to estimate the parameters for a coalescing binary system for signals detected by LIGO antennas. While we have not yet optimized this procedure for speed, we have shown that within just a few hours of running time on a $500MHz$ pc we can generate kernel densities for the parameters. The MCMC can replace a systematic march through a grid of templates in parameter space, and instead make a probabilistic random walk that is helped by the weighting of the product of the likelihood and the *a priori* distributions of the parameters.

We have only concentrated on using our MCMC procedure for estimating signals in data sets where a signal is assumed to exist. Using MCMC techniques as a method to find signals from the continuous output of the LIGO detectors is another research topic that is not covered here. A question that must be answered is how efficiently can the MCMC method identify signals, and how often does it miss them.

A great advantage of MCMC methods is that the calculational time does not scale exponentially with parameter number, but in fact scales almost linearly. Applications of state-space modelling in finance, such as stochastic volatility models applied to time series of daily exchange rates or returns of stock exchange indices, easily have 1000 - 5000 parameters; specially tailored MCMC algorithms can effectively and efficiently sample the phase space [27,37–39]. The number of parameters for coalescing binary signals will grow when, for example, the spin of the compact objects is included. We will also eventually expand our MCMC study to the problem of examining the signals detected by two or more interferometers simultaneously. In addition to the parameters for pertaining to the coalescing binary, there will be the delay in arrival times between the detectors and the polarization sensitivities that can then infer the location in the sky of the source [40]. MCMC methods offer great promise for parameter estimation with coalescing binary signals, especially as parameter numbers increase.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation Grant PHY-0071327, Carleton College, The Royal Society of New Zealand Marsden Fund, and the University of Auckland Research Committee.

REFERENCES

* Authors' email addresses: nchriste@carleton.edu, meyer@stat.auckland.ac.nz.

[1] A.Abramovici, W.E.Althouse, R.W.P.Drever, Y. Gürsel, S.Kawamura, F.J. Raab, D. Shoemaker, L. Sievers, R.E. Spero, K.S. Thorne, R.E. Vogt, R. Weiss, S.E. Whitcomb, M.E. Zucker, Science **256**, 325 (1992).

[2] C. Bradaschia *et al.*, Nucl. Instrum. Methods A **289**, 518 (1990).

[3] B. Caron *et al.*, Nuclear Physics; Suppl. **48**, 107 (1996).

[4] K.S. Thorne, in *300 Years of Gravitation*, edited by S.W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1987), p. 330.

[5] L.S. Finn and D. Chernoff, Phys. Rev. D **47**, 2198 (1993).

[6] B.F. Schutz, Nature **323**, 310 (1996).

[7] C. Cutler and E.E. Flanagan, Phys. Rev. D **49**, 2658 (1994).

[8] L. S. Finn, Phys. Rev. D **53**, 2878 (1996).

[9] C. Cutler et al., Phys. Rev. Lett. **70**, 2984 (1993).

[10] E.E. Flanagan and S.A. Hughes, Phys. Rev. D **57**, 4535 (1998).

[11] E.E. Flanagan and S.A. Hughes, Phys. Rev. D **57**, 4566 (1998).

[12] P.C. Gregory and T. J. Loredo, Astrophys. J. **398**, 146 (1992).

[13] L.S. Finn, 1997, gr-qc/9709077.

[14] L.S. Finn, Phys. Rev. D **46**, 5236 (1992).

[15] F. Nicholson and A. Vecchio, Phys. Rev. D **57**,4588 (1998).

[16] W.R. Gilks and S. Richardson and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).

[17] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[18] S. Geman and D. Geman, IEEE Trans. Patt. Anal. Mach. Intell. **6**, 721 (1984).

[19] D.J. Spiegelhalter, A. Thomas, N. Best, W. R. Gilks, *BUGS 0.5 Bayesian Inference Using Gibbs Sampling Manual (Version ii)* (Cambridge UK: MRC Biostatistics Unit, Cambridge, 1996).

[20] N.L. Christensen and R. Meyer, Phys. Rev. D **58**, 082001 (1998).

[21] L. Devroye, *Non-Uniform Random Variate Generation* (Springer, New York, 1986).

[22] W. R. Gilks and P. Wild, Appl. Statist. **41**, 169 (1992).

[23] W.R. Gilks, N.G. Best and K. Chan, Appl. Statist. **44**, 445 (1995).

[24] A. Thomas and D.J. Spiegelhalter and W.R. Gilks, in *Bayesian Statistics 4*, edited by J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, (Oxford University Press, Oxfird, 1992), pp. 837-842.

[25] W.R. Gilks, A. Thomas, D.J. Spiegelhalter, The Statistician **43**, 169 (1994).

[26] R. Gentleman, Chance **10**, 48 (1997).

[27] R. Meyer, in *Data Analysis: Scientific Modeling and Practical Applications*, edited by W. Gaul, O. Opitz and M. Schader, (Springer, Berlin, 2000), pp. 259-271.

[28] N. G. Best and M. K. Cowles and S. K. Vines, *CODA Manual Version 0.30* (Cambridge, UK: MRC Biostatistics Unit, Cambridge, 1995).

[29] L. Blanchet, Phys. Rev. D **54**, 1417 (1996).

[30] L. Blanchet, B. R. Iyer, C. M. Will, A. G. Wiseman, Class.Quant.Grav. **13**, 575 (1996).

[31] T. Tanaka, H. Tagoshi, Phys. Rev. D **62**, 082001 (2000).

[32] L.S. Finn and E. Daw, http://fiji/nirvana.psu.edu (2001).

[33] N. Christensen and R. Meyer, http://physics.carleton.edu/Faculty/nelsonhome.html (2001).

[34] W.R. Gilks, G.O. Roberts, in *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson and D.J. Spiegelhalter, (Chapman and Hall, London, 1996), pp. 89-114.

[35] C. Jennison, J. R. Statist. Soc. B. **55**, 54 (1993).

[36] B. S. Sathyaprakesh, Class.Quant.Grav. **17**, L157 (2000).

[37] E. Jacquier, N.G. Polson, P.E. Rossi, J. Bus. Econ. Stat. **12**, 371 (1994).

[38] R. Meyer and J. Yu, Econometrics J. **3**, 198 (2000).

[39] S. Kim, N. Shephard, S. Chib, Rev. Econ. Studies **65**, 361 (1998).

[40] W.G. Anderson, P.R. Brady, J.D.E. Creighton, E.E. Flanagan, Physical Rev. D **6**3, 042003 (2001).

Figure Captions

Fig. 1 Estimate of the total mass, m_t, versus SNR. The actual total mass for the signal is $m_t = 4.9M_\odot$. Error bars correspond to the 2.5 and 97.5 percentile of the posterior distribution of m_t (which gives a 95% posterior credibility interval for m_t).

Fig. 2 Estimate of the mass ratio, η, versus SNR. The actual mass ratio for the signal is $\eta = 0.2041$. Error bars correspond to the 2.5 and 97.5 percentile of the posterior distribution of η (which gives a 95% posterior credibility interval for η).

Fig. 3 Estimate of the (rescaled) amplitude of the gravity wave, N, versus SNR. The amplitude of the inferred signal varies linearly with the SNR, as it should. Error bars correspond to the 2.5 and 97.5 percentile of the posterior distribution of N (which gives a 95% posterior credibility interval for N).

Fig. 4 The kernel densities for the parameters (m_t, η, N, t_c, and φ_0), generated from the Gibbs sampler with $SNR = 4.4$.

Fig. 5 Example of the operation of the Gibbs sampler, with trace plots for the parameters displayed, with $SNR = 4.4$ and the simulated annealing pseudo-temperature of $T = 100$. One can see how the chain *burns-in* and achieves convergence.

Fig. 6 Estimate of the total mass, m_t, versus SNR. The actual total mass for the signal is $m_t = 17M_\odot$. Error bars correspond to the 2.5 and 97.5 percentile of the posterior distribution of m_t (which gives a 95% posterior credibility interval for m_t).

Fig. 7 Estimate of the mass ratio, η, versus SNR. The actual mass ratio for the signal is $\eta = 0.2491$. Error bars correspond to the 2.5 and 97.5 percentile of the posterior distribution of η (which gives a 95% posterior credibility interval for η).

Fig. 8 Estimate of the (rescaled) amplitude of the gravity wave, N, versus SNR. The amplitude of the inferred signal varies linearly with the SNR, as it should. Error bars correspond to the 2.5 and 97.5 percentile of the posterior distribution of N (which gives a 95% posterior credibility interval for N).

Fig. 9 The kernel densities for the parameters (m_t, η, N, t_c, and φ_0), generated from the Gibbs sampler with $SNR = 4.4$.

Fig. 10 Example of the operation of the Gibbs sampler, with trace plots for the parameters displayed, with $SNR = 4.4$ and the simulated annealing pseudo-temperature of $T = 100$. One can see how the chain *burns-in* and achieves convergence.